

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平9-160589

(43)公開日 平成9年(1997)6月20日

(51)Int.Cl. <sup>8</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 7 1		G 1 0 L 3/00	5 7 1 K
	5 3 1			5 3 1 L
G 0 6 F 15/00	3 3 0		G 0 6 F 15/00	3 3 0 F
G 0 6 T 7/00			15/62	4 6 5 U

審査請求 未請求 請求項の数13 O L (全 24 頁)

(21)出願番号 特願平7-317942

(22)出願日 平成7年(1995)12月6日

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 小原 和昭

大阪府門真市大字門真1006番地 松下電器  
産業株式会社内

(72)発明者 渡辺 辰巳

大阪府門真市大字門真1006番地 松下電器  
産業株式会社内

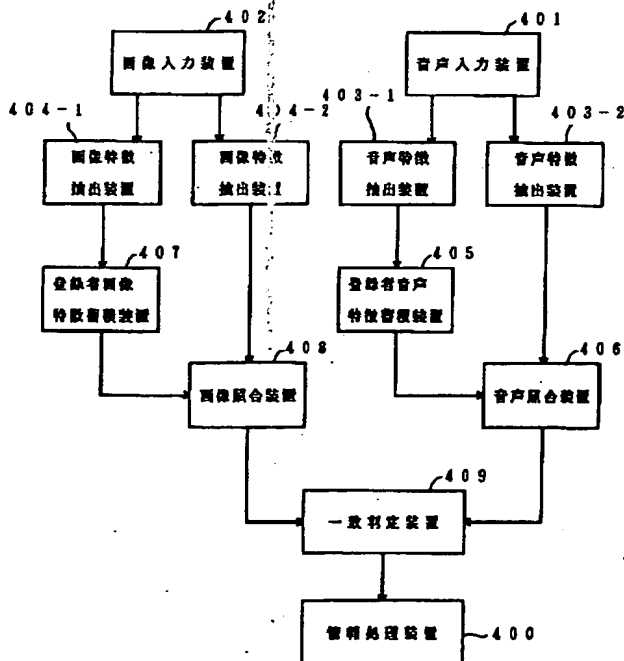
(74)代理人 弁理士 岡田 和秀

## (54)【発明の名称】 利用者照合装置

## (57)【要約】

【課題】 キーボードからパスワードを入力する場合に比べて、利用者照合をより厳密に行うことができ、不正にパスワードを取得した利用者が情報処理装置を不正に利用することを防止することができる利用者照合装置を提供する。

【解決手段】 利用者の音声を蓄積する手段405と、利用者の画像を蓄積する手段407と、現在の利用者の画像を入力する手段402と、前記入入力画像から利用者の画像特徴量を抽出する手段404と、現在の利用者の音声を入力する手段401と、前記入入力音声から利用者の音声特徴量を抽出する手段403と、前記入入力画像と入力音声を前記蓄積されている画像・音声と照合する手段408、406と、前記照合に用いる音声または画像を限定する手段と、前記照合結果に基づき現在の利用者が登録されている利用者であるか否かを判断する手段409から構成される。



## 【特許請求の範囲】

【請求項1】 情報処理装置の利用許可された利用者の画像の特徴量と音声の特徴量を蓄積しておき、入力した現在の利用者の画像の特徴量を蓄積されている画像の特徴量と照合するとともに、入力した現在の利用者の音声の特徴量を蓄積されている音声の特徴量と照合し、画像による照合結果と音声による照合結果とが一致したときに現在の利用者が登録されている利用者であると判断するように構成してあることを特徴とする利用者照合装置。

【請求項2】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記入力画像に含まれる画像特徴量を予め蓄積してある画像特徴量と照合する手段と、前記画像照合結果により前記蓄積手段に蓄積してある利用者を照合するために用いる音声の候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項3】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記入力画像に含まれる画像特徴量を予め蓄積してある画像特徴量と照合する手段と、前記画像照合結果と入力された利用者の個人性を反映する音声特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項4】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記入力音声に含まれる音声特徴量を予め蓄積してある音声特徴量と照合する手段と、前記音声照合結果により前記蓄積してある利用者の照合に用いる画像候補を限定する手段と、前記限定された画像候補の特徴量と入力された画像特徴量とを照合する手段と、前記画像による照合結果と

前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項5】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、入力された利用者の個人性を反映する音声特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、入力された利用者の個人性を反映する画像特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる画像候補を限定する手段と、前記限定された画像候補の特徴量と入力された画像の特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項6】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、入力された利用者の個人性を反映する音声特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、入力された利用者の個人性を反映する画像特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる画像候補を限定する手段と、前記音声照合結果により同様に蓄積してある利用者を照合するために用いる画像候補を限定する手段と、前記限定された画像候補の特徴量と入力された画像の特徴量とを照合する手段と、前記音声による照合結果と前記画像による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項7】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、入力された利用者の個人性を反映する画像特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる画像候補を限定する手段と、前記限定された画像候補

の特徴量と入力された画像の特徴量とを照合する手段と、入力された利用者の個人性を反映する音声特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記画像照合結果により同様に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項 8】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記入力音声に含まれる音声特徴量を予め蓄積してある音声特徴量と照合する手段と、前記入力音声から現在の利用者を照合したときの音声信頼度を求める手段と、前記音声照合結果により前記蓄積してある利用者の照合に用いる画像を限定する手段と、前記限定された画像特徴量と入力された画像特徴量とを照合する手段と、前記画像照合結果から現在の利用者を照合したときの画像信頼度を求める手段と、前記音声からの信頼度と前記画像からの信頼度の各々が予め設定した値よりも高いときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項 9】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、前記入力した画像特徴量と予め蓄積してある画像特徴量とを照合する手段と、前記画像照合結果の信頼度を求める手段と、前記画像照合結果と信頼度に基づき音声照合に用いるための登録利用者の数と候補を限定する手段と、前記限定された利用者の音声特徴量と入力された音声特徴量とを照合する手段と、前記音声照合結果の信頼度を求める手段と、前記画像から求めた信頼度と前記音声から求めた信頼度の各々が予め設定した値よりも高いときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項 10】 登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、現在の利用者の音声を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、前記入力した画像特徴量と予め蓄積してある画像特徴量とを照合する手段と、前記画

像照合結果の信頼度を求める手段と、前記画像照合結果と信頼度に基づき音声照合に用いるための登録利用者の候補を限定する手段と、前記画像信頼度が予め設定した基準値より高かったときに現在の利用者が登録利用者であると判断する手段と、前記画像信頼度が予め設定した基準値よりも低かったときに前記音声入力手段から音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記画像照合結果と前記音声の個人性情報を用いて限定された音声特徴量と入力された音声特徴量とを照合する手段と、前記音声照合結果の信頼度を求める手段と、前記音声の信頼度と前記画像の信頼度の各々が予め設定した値よりも高かったときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴とする利用者照合装置。

【請求項 11】 利用者の音声特徴量を抽出する手段が、利用者の音声のピッチの平均値を求める手段と、ピッチの標準偏差を求める手段と、平均音声パワーを求める手段と、平均音声パワーの標準偏差を求める手段と、音声パワースペクトラムの時系列情報を求める手段とからなる請求項 1 から請求項 10 までのいずれかに記載の利用者照合装置。

【請求項 12】 利用者の画像特徴量を抽出する手段が、利用者の顔画像の輝度情報を求める手段と、輝度情報の位置微分情報を求める手段と、利用者の顔画像の輪郭情報を求める手段と、利用者の顔画像の色情報を求める手段と、前記画像の輝度情報、輝度情報の位置微分情報、輪郭情報、色情報の重みづけを行う手段とからなる請求項 1 から請求項 10 までのいずれかに記載の利用者照合装置。

【請求項 13】 利用者の画像特徴量を抽出する手段が、画像から利用者の身長を推定する手段と、利用者の顔画像の色情報を求める手段と、前記顔画像の輝度情報、輝度情報の位置微分情報、輪郭情報とからなる請求項 1 から請求項 10 までのいずれかに記載の利用者照合装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は利用者照合装置に関するものであり、特に音声と画像の個人性特徴量を利用して効率的な利用者照合を行う利用者照合装置に関する。

【0002】

【従来の技術】 従来の利用者照合装置は、一般的に例えば図 11 に示したように構成されている。図 11 において、1101 は情報入力するための情報入力手段で例えばキーボード、1102 は情報処理装置で例えばコンピュータ、1103 は入力された情報を表示するための情報表示手段、1104 は他の情報処理装置と接続するためのネットワーク装置である。

【0003】 以上のように構成された従来の利用者照合

装置においては、情報入力開始に当たって情報処理装置を現在の利用者（これから利用しようとする利用者）が予め利用許可された利用者であるか否かを、通常、利用者がキーボードからパスワード（暗唱番号）を入力することにより確認して、利用許可された利用者であると判断したときに限って情報の入力が可能になるようにする。

#### 【0004】

【発明が解決しようとする課題】しかしながら、以上のように構成された従来の利用者照合装置においては、不特定多数の利用者が利用できるような環境で情報処理装置が利用されるような場合には、何らかの手段によって、利用許可された以外の者が不正にパスワードを取得してキーボードからパスワードを入力することにより情報処理装置が不正に利用されてしまうという問題があった。

【0005】本発明はかかる点に鑑み、不正にパスワードを取得した利用者が情報処理装置を不正に利用することを防止し、情報処理装置の利用者の照合をより厳密に行うことのできる利用者照合装置を提供することを目的とする。

#### 【0006】

【課題を解決するための手段】本発明に係る利用者照合装置は、情報処理装置の利用を許可された利用者の画像の特徴量と音声の特徴量を予め蓄積しておき、現在の利用者の画像と音声を入力し、それぞれの特徴量を抽出し、抽出した画像特徴量および音声特徴量をそれぞれ蓄積されている画像特徴量および音声特徴量と照合して、画像による照合結果と音声による照合結果とが一致したときに現在の利用者が登録されている利用者であると判断するように構成してあることを特徴としている。利用者の画像の特徴量と音声の特徴量を用いて照合するので、情報処理装置の利用者の照合をより厳密に行うことができ、不正にパスワードを取得した利用者が情報処理装置を不正に利用することを防止する。

#### 【0007】

【発明の実施の形態】本発明に係る請求項1の利用者照合装置は、情報処理装置の利用許可された利用者の画像の特徴量と音声の特徴量を蓄積しておき、入力した現在の利用者の画像の特徴量を蓄積されている画像の特徴量と照合するとともに、入力した現在の利用者の音声の特徴量を蓄積されている音声の特徴量と照合し、画像による照合結果と音声による照合結果とが一致したときに現在の利用者が登録されている利用者であると判断するように構成してあることを特徴としている。利用者の画像の特徴量と音声の特徴量を用いて照合するので、情報処理装置の利用者の照合をより厳密に行うことができ、不正にパスワードを取得した利用者が情報処理装置を不正に利用することを防止する。

【0008】本発明に係る請求項2の利用者照合装置

は、登録利用者の音声と蓄積する手段と、登録利用者の画像と蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記入力画像に含まれる画像特徴量を予め蓄積してある画像特徴量と照合する手段と、前記画像照合結果により前記蓄積手段に蓄積してある利用者を照合するために用いる音声の候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴としている。音声照合するに際して、予め、画像による照合結果に基づいて音声候補を限定してあるから、照合すべき数が大きいとき、効率的な照合が可能となる。

【0009】本発明に係る請求項3の利用者照合装置は、登録利用者の音声と蓄積する手段と、登録利用者の画像と蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記入力画像に含まれる画像特徴量を予め蓄積してある画像特徴量と照合する手段と、前記画像照合結果と入力された利用者の個人性を反映する音声特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴としている。まず現在の利用者の画像情報を用いて利用者の照合を行い、その結果と音声の大局的な特徴量の類似性を用いて照合すべき音声を限定することによって、照合すべき画像と音声の数が大きいとき、効率的な照合が可能となる。

【0010】本発明に係る請求項4の利用者照合装置は、登録利用者の音声と蓄積する手段と、登録利用者の画像と蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記入力音声に含まれる音声特徴量を予め蓄積してある音声特徴量と照合する手段と、前記音声照合結果により前記蓄積してある利用者の照合に用いる画像候補を限定する手段と、前記限定された画像候補の特徴量と入力された画像特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者

であると判断する手段を備えたことを特徴としている。  
まず現在の利用者の音声情報を用いて利用者の照合を行い、その照合結果を用いて照合すべき画像を大幅に限定することによって、音声に比べて照合時間のかかる画像の照合時間を大きく削減することができ、登録者数が大きな情報処理装置を利用する場合に効率的な照合が可能となる。

【0011】本発明に係る請求項5の利用者照合装置は、登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、入力された利用者の個人性を反映する音声特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、入力された利用者の個人性を反映する画像特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる画像候補を限定する手段と、前記限定された画像候補の特徴量と入力された画像の特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴としている。利用者の音声の大局的な情報と画像の大局的な情報とを用いて照合候補を選択することによって、利用者照合時間を大きく削減することができ、利用者数が大きな情報処理装置を利用する場合に効率的な照合が可能となる。

【0012】本発明に係る請求項6の利用者照合装置は、登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、入力された利用者の個人性を反映する音声特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、入力された利用者の個人性を反映する画像特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる画像候補を限定する手段と、前記音声照合結果により同様に蓄積してある利用者を照合するために用いる画像候補を限定する手段と、前記限定された画像候補の特徴量と入力された画像の特徴量とを照合する手段と、前記音声による照合結果と前記画像による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴としている。音声照合結果を用いて照合すべき画像をさらに限定するので、請求項5の

場合よりもより効率的な照合を行うことができる。

【0013】本発明に係る請求項7の利用者照合装置は、登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、入力された利用者の個人性を反映する画像特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる画像候補を限定する手段と、前記限定された画像候補の特徴量と入力された画像の特徴量とを照合する手段と、入力された利用者の個人性を反映する音声特徴量の一部を用いて前記蓄積手段に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記画像照合結果により同様に蓄積してある利用者を照合するために用いる音声候補を限定する手段と、前記限定された音声候補の特徴量と入力された音声の特徴量とを照合する手段と、前記画像による照合結果と前記音声による照合結果が一致したときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴としている。画像照合結果を用いて照合すべき画像をさらに限定するので、請求項5の場合よりもより効率的な照合を行うことができる。

【0014】本発明に係る請求項8の利用者照合装置は、登録利用者の音声と蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記入力音声に含まれる音声特徴量を予め蓄積してある音声特徴量と照合する手段と、前記入力音声から現在の利用者を照合したときの音声信頼度を求める手段と、前記音声照合結果により前記蓄積してある利用者の照合に用いる画像を限定する手段と、前記限定された画像特徴量と入力された画像特徴量とを照合する手段と、前記画像照合結果から現在の利用者を照合したときの画像信頼度を求める手段と、前記音声からの信頼度と前記画像からの信頼度の各々が予め設定した値よりも高いときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴としている。入力された音声と画像を用いて利用者を判定すると同時に、音声と画像の信頼度を考慮することでより信頼度の高い確実な利用者照合判定を行うことができる。

【0015】本発明に係る請求項9の利用者照合装置は、登録利用者の音声と蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、現在の利用者の画像を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、前記入力した画像特徴

量と予め蓄積してある画像特徴量とを照合する手段と、前記画像照合結果の信頼度を求める手段と、前記画像照合結果と信頼度に基づき音声照合に用いるための登録利用者の数と候補を限定する手段と、前記限定された利用者の音声特徴量と入力された音声特徴量とを照合する手段と、前記音声照合結果の信頼度を求める手段と、前記画像から求めた信頼度と前記音声から求めた信頼度の各々が予め設定した値よりも高いときに現在の利用者が登録されている利用者であると判断する手段を備えたことを特徴としている。入力された画像と音声を用いて利用者を判定すると同時に、音声と画像の信頼度を考慮することでより信頼度の高い確実な利用者照合判定を行うことができる。

【0016】本発明に係る請求項10の利用者照合装置は、登録利用者の音声を蓄積する手段と、登録利用者の画像を蓄積する手段と、現在の利用者の画像を入力する手段と、現在の利用者の音声を入力する手段と、前記入力画像から利用者の個人性を反映する画像特徴量を抽出する手段と、前記入力した画像特徴量と予め蓄積してある画像特徴量とを照合する手段と、前記画像照合結果の信頼度を求める手段と、前記画像照合結果と信頼度に基づき音声照合に用いるための登録利用者の候補を限定する手段と、前記画像信頼度が予め設定した基準値より高かったときに現在の利用者が登録利用者であると判断する手段と、前記画像信頼度が予め設定した基準値よりも低かったときに前記音声入力手段から音声を入力する手段と、前記入力音声から利用者の個人性を反映する音声特徴量を抽出する手段と、前記画像照合結果と前記音声の個人性情報を用いて限定された音声特徴量と入力された音声特徴量とを照合する手段と、前記音声照合結果の信頼度を求める手段と、前記音声の信頼度と前記画像の信頼度の各々が予め設定した値よりも高かったときに現在の利用者が登録利用者であると判断する手段を備えたことを特徴としている。まず画像を用いて利用者を照合し、そのときの画像信頼度が基準値よりも高かった場合には音声を使った照合は行わずに情報処理装置の利用を許可し、画像信頼度が基準値よりも低かった場合には音声を用いた照合も行うことで、より効率的で確かな利用者照合判定を行うことができる。

【0017】本発明に係る請求項11の利用者照合装置は、上記請求項1から請求項10までのいずれかにおいて、利用者の音声特徴量を抽出する手段が、利用者の音声のピッチの平均値を求める手段と、ピッチの標準偏差を求める手段と、平均音声パワーを求める手段と、平均音声パワーの標準偏差を求める手段と、音声パワースペクトラムの時系列情報を求める手段とからなるものである。高精度な照合を行うことができる。

【0018】本発明に係る請求項12の利用者照合装置は、上記請求項1から請求項10までのいずれかにおいて、利用者の画像特徴量を抽出する手段が、利用者の顔

画像の輝度情報を求める手段と、輝度情報の位置微分情報を求める手段と、利用者の顔画像の輪郭情報を求める手段と、利用者の顔画像の色情報を求める手段と、前記画像の輝度情報、輝度情報の位置微分情報、輪郭情報、色情報の重みづけを行う手段とからなるものである。高精度な照合を行うことができる。

【0019】本発明に係る請求項13の利用者照合装置は、上記請求項1から請求項10までのいずれかにおいて、利用者の画像特徴量を抽出する手段が、画像から利用者の身長を推定する手段と、利用者の顔画像の色情報を求める手段と、前記顔画像の輝度情報、輝度情報の位置微分情報、輪郭情報とからなるものである。高精度な照合を行うことができる。

【0020】（実施の形態1）以下、本発明の実施の形態1について図面を参照して説明する。

【0021】図1は本発明の実施の形態1における利用者照合装置の構成図を示すものである。図1において、100は情報処理装置、101は情報処理装置100をこれから利用しようとする利用者（現在の利用者）の音声を入力するための音声入力装置、102は情報処理装置100の現在の利用者の顔画像を入力するための画像入力装置、103は音声入力装置101で入力された現在の利用者の音声の特徴量を取り出すための音声特徴抽出装置、104は画像入力装置102で入力された現在の利用者の画像の特徴量を取り出すための画像特徴抽出装置、105は予め登録されている利用者の音声の特徴量を蓄積している登録者音声特徴蓄積装置、106は音声特徴抽出装置103で取り出された現在の利用者の音声の特徴量と登録者音声特徴蓄積装置105に予め登録されている利用者の音声の特徴量を照合するための音声照合装置である。この音声照合装置106は、画像照合装置108での画像照合の結果により登録者音声特徴蓄積装置105で限定されたN個（Nは正の整数）の音声候補の特徴量を音声特徴抽出装置103が抽出した音声の特徴量と照合するように構成されている。107は予め登録されている利用者の画像の特徴量を蓄積している登録者画像特徴蓄積装置、108は画像特徴抽出装置104で取り出された現在の利用者の画像の特徴量と登録者画像特徴蓄積装置107に予め登録されている利用者の画像の特徴量を照合するための画像照合装置、109は画像照合装置108で照合された利用者と音声照合装置106で照合された利用者とが一致しているか否かを判定する一致判定装置である。

【0022】以上のように構成された実施の形態1の利用者照合装置について、以下にその動作を説明する。情報処理装置100の現在の利用者は利用に先立ち利用者の音声と顔画像をそれぞれ音声入力装置101と画像入力装置102から入力する。

【0023】入力された画像は画像特徴抽出装置104に入力され、現在の利用者の画像に含まれる個人性情報

を抽出する。画像に含まれる個人性情報としては、顔画像の輝度情報、輝度情報の位置微分情報、顔画像の色情報を用いる。現在の利用者の画像が予め情報処理装置100に登録されている利用者であるか否かの照合については、画像特徴抽出装置104で抽出した特徴量と登録者画像特徴蓄積装置107の特徴量とが画像照合装置108で照合され、利用者が予め登録されているか否かを照合する。

【0024】一方、入力された音声は音声特徴抽出装置103に入力され、現在の利用者の音声に含まれる個人性情報を抽出する。ここで抽出している個人性情報としては、音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差、音声パワースペクトラムの時系列情報等を用いる。これらの特徴量は公知の手法

(例えばDFT分析、LPC分析、フィルタ分析等)を用いて容易に求めることが可能である。音声特徴抽出装置103で抽出された音声の特徴量は登録者音声特徴蓄積装置105に蓄積されている予め登録されている利用者の音声の特徴量と音声照合装置106で照合され、利用者が予め登録されているか否かを照合する。ここで照合に用いられる蓄積音声は、画像照合の結果から求めたN個(Nは正の整数)の利用者の候補の音声に限定されている。すなわち、音声照合装置106は画像照合の結果により登録者音声特徴蓄積装置105で限定されたN個の音声候補の特徴量を音声特徴抽出装置103が抽出した音声の特徴量と照合する。音声照合装置106での特徴量の照合は、公知のHMM(Hidden Markov Model)、DTW(Dynamic Time Warping)、LVQ(Learning Vector Quantizer)などのパターン照合手法を用いて行うことができる。一致判定装置109は画像照合装置108による照合結果と音声照合装置106による照合結果とが一致したときのみ「1」を出力し、不一致のときは「0」を出力する。一致判定装置109が「1」を出力したときのみ、現在の利用者が登録されている利用者であると判断して、現在の利用者に対して情報処理装置100の利用を許可する。

【0025】以上のように本実施の形態によれば、利用者に特有な画像情報と音声情報を用いて利用者の照合を行うことによって、従来のパスワード(暗唱番号)をキーボードから入力するだけでは避けることのできない情報処理装置の不正利用を防ぐことができる。そして、音声照合するに際して、予め、画像による照合結果に基づいて音声候補を限定してあるから、照合すべき数が大きいとき、効率的な照合が可能となる。

【0026】(実施の形態2)以下、本発明の実施の形態2について図面を参照して説明する。

【0027】図2は本発明の実施の形態2における利用者照合装置の構成図を示すものである。図2において、200は情報処理装置、201は情報処理装置200に現在の利用者の音声を入力するための音声入力装置、2

02は情報処理装置200の現在の利用者の顔画像を入力するための画像入力装置、203-1は音声入力装置201で入力された現在の利用者の音声の平均的なピッチ等の大局的な特徴量を取り出すための音声特徴抽出装置、203-2は音声入力装置201で入力された現在の利用者の音声のスペクトル特徴を反映する特徴量を取り出すための音声特徴抽出装置、204は画像入力装置202で入力された現在の利用者の画像の特徴量を取り出すための画像特徴抽出装置、205は予め登録されている利用者の音声の特徴量を蓄積している登録者音声特徴蓄積装置、206は音声特徴抽出装置203で取り出された現在の利用者の音声の特徴量と予め登録されている利用者の音声の特徴量を照合するための音声照合装置、207は予め登録されている利用者の画像の特徴量を蓄積している登録者画像特徴蓄積装置、208は画像特徴抽出装置204で取り出された現在の利用者の画像の特徴量と予め登録されている利用者の画像の特徴量を照合するための画像照合装置、209は画像照合装置208で照合された利用者と音声照合装置206で照合された利用者とは一致しているか否かを判定する一致判定装置である。音声特徴抽出装置203-1による大局的な音声の特徴量に基づいて登録者音声特徴蓄積装置205での音声候補を限定するように構成されている。また、画像照合装置208によるN個の画像候補に基づいても登録者音声特徴蓄積装置205での音声候補を限定するように構成されている。

【0028】以上のように構成された実施の形態2の利用者照合装置について、以下にその動作を説明する。情報処理装置200の現在の利用者は利用に先立ち利用者の音声と顔画像を入力する。画像入力装置202によって入力された画像は画像特徴抽出装置204に入力され、利用者の画像に含まれる個人性情報を抽出する。画像に含まれる個人性情報としては、顔画像の輝度情報、輝度情報の位置微分情報、顔画像の色情報を用いる。画像特徴抽出装置204で取り出した画像特徴量と登録者画像特徴蓄積装置207に蓄積された画像特徴量とが画像照合装置208で照合され、現在の利用者に最も近い登録利用者からN番目(Nは正の整数)までの候補を出力する。

【0029】一方、音声入力装置201から入力された音声は音声特徴抽出装置203-1、203-2に入力され、利用者の音声に含まれる個人性情報を抽出する。音声特徴抽出装置203-1では、音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差の個人性情報の大局的な特徴量を求める。音声特徴抽出装置203-2では、音声パワースペクトラムの時系列情報を求める。これらの特徴量は公知の手法(例えばDFT分析、LPC分析、フィルタ分析等)を用いて容易に求めることが可能である。音声特徴抽出装置203で求められた音声の特徴量は登録者音声特徴蓄積装置20



5に予め登録されている利用者の音声の特徴量と音声照合装置206で照合され、利用者が予め登録されているか否かを照合する。ここで照合に用いられる蓄積音声は、画像照合の結果から求めたN個の利用者の候補と、音声特徴抽出装置203-1で求めた大局的な音声特徴量の類似している登録利用者に限定されている。音声照合装置206での特徴量の照合は公知のHMM、DTW、LVQなどのパターン照合手法を用いて行うことができる。一致判定装置209は画像照合装置208による照合結果と音声照合装置206による照合結果とが一致したときのみ「1」を出力し、不一致のときは「0」を出力する。一致判定装置209が「1」を出力したときのみ、現在の利用者が登録されている利用者であると判断して、現在の利用者に対して情報処理装置200の利用を許可する。

【0030】以上のように本実施の形態によれば、まず現在の利用者の画像情報を用いて利用者の照合を行い、その結果と音声の大局的な特徴量の類似性とを用いて照合すべき音声を限定することによって、照合すべき画像と音声の数が大きいとき、効率的な照合が可能となるとともに、従来のパスワードをキーボードから入力するだけでは避けることのできない情報処理装置の不正利用を防ぐことができる。

【0031】（実施の形態3）以下、本発明の実施の形態3について図面を参照して説明する。

【0032】図3は本発明の実施の形態3における利用者照合装置の構成図を示すものである。図3において、300は情報処理装置、301は情報処理装置300に現在の利用者の音声を入力するための音声入力装置、302は情報処理装置300の現在の利用者の顔画像を入力するための画像入力装置、303-1は音声入力装置301で入力された現在の利用者の音声の平均的なピッチ等の大局的な特徴量を取り出すための音声特徴抽出装置、303-2は音声入力装置301で入力された現在の利用者の音声のスペクトル特徴を反映する特徴量を取り出すための音声特徴抽出装置、304は画像入力装置302で入力された現在の利用者の画像の特徴量を取り出すための画像特徴抽出装置、305は予め登録されている利用者の音声の特徴量を蓄積している登録者音声特徴蓄積装置、306は音声特徴抽出装置303で取り出された現在の利用者の音声の特徴量と予め登録されている利用者の音声の特徴量を照合するための音声照合装置、307は予め登録されている利用者の画像の特徴量を蓄積している登録者画像特徴蓄積装置、308は画像特徴抽出装置304で取り出された現在の利用者の画像の特徴量と予め登録されている利用者の画像の特徴量を照合するための画像照合装置、309は画像照合装置308で照合された利用者と音声照合装置306で照合された利用者とが一致しているか否かを判定する一致判定装置である。音声特徴抽出装置303-1による大局的

な音声の特徴量に基づいて登録者音声特徴蓄積装置305での音声候補を限定するように構成されている。また、音声照合装置306によるN個の音声候補に基づいて登録者画像特徴蓄積装置307での画像候補を限定するように構成されている。

【0033】以上のように構成された実施の形態3の利用者照合装置について、以下にその動作を説明する。情報処理装置300の現在の利用者は利用に先立ち利用者の音声と顔画像を入力する。音声入力装置301から入力された音声は音声特徴抽出装置303-1、303-2に入力され利用者の音声に含まれる個人性情報を抽出する。音声特徴抽出装置303-1では、音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差の個人性情報の大局的な特徴量を求める。音声特徴抽出装置303-2では、音声パワースペクトラムの時系列情報等を求める。これらの特徴量は公知の手法

（例えばDFT分析、LPC分析、フィルタ分析等）を用いて容易に求めることが可能である。音声特徴抽出装置303で求められた音声の特徴量は登録者音声特徴蓄積装置305に予め登録されている利用者の音声の特徴量と音声照合装置306によって照合され、利用者が予め登録されているか否かが判断される。照合を効率的に行うために本実施の形態では、音声特徴抽出装置303-1で抽出された音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差の個人性情報の大局的な特徴量の類似した登録者音声特徴蓄積装置305の音声に限定して音声照合を行っている。音声照合装置306での特徴量の照合は公知のHMM、DTW、LVQなどのパターン照合手法を用いて行うことができる。

【0034】一方、画像入力装置302によって入力された現在の利用者の画像は画像特徴抽出装置304で特徴抽出された後、画像照合装置308に入力され、登録者画像特徴蓄積装置307に蓄積された画像特徴量と照合され、現在の利用者が予め登録されている利用者であるか否かが判断される。本実施の形態では、画像に含まれる個人性情報として、顔画像の輝度情報、輝度情報の位置微分情報、顔画像の色情報を用いる。現在の利用者の画像が予め情報処理装置300に登録されている利用者であるか否かは、画像特徴抽出装置304で取り出した特徴量と登録者画像特徴蓄積装置307の特徴量とを画像照合装置308で照合する。照合すべき登録蓄積画像の候補は本実施の形態では、先の音声照合装置306での照合結果によって限定された候補を用いている。この画像照合装置308では現在の利用者の入力画像と最も類似した蓄積画像が出力される。一致判定装置309は画像による照合結果と前記音声による照合結果が一致したときに「1」を出力し、不一致のときは「0」を出力する。一致判定装置309が「1」を出力したときのみ、現在の利用者が登録されている利用者であると判断して、現在の利用者に対して情報処理装置300を利用



することを許可する。

【0035】以上のように本実施の形態によれば、まず現在の利用者の音声情報を用いて利用者の照合を行い、その照合結果を用いて照合すべき画像を大幅に限定することによって、音声に比べて照合時間のかかる画像の照合時間を大きく削減することができ、登録者数が大きな情報処理装置を利用する場合に効率的な照合が可能となるとともに、従来のパスワードをキーボードから入力するだけでは避けることのできない情報処理装置の不正利用を防ぐことができる。

【0036】(実施の形態4)以下、本発明の実施の形態4について図面を参照して説明する。

【0037】図4は本発明の実施の形態4における利用者照合装置の構成図を示すものである。図8において、400は情報処理装置、401は情報処理装置400に現在の利用者の音声を入力するための音声入力装置、402は情報処理装置400に現在の利用者の画像を入力するための画像入力装置、403-1は音声入力装置401で入力された現在の利用者の音声の平均的なピッチ等の大局的な特徴量を取り出すための音声特徴抽出装置、403-2は音声入力装置401で入力された現在の利用者の音声のスペクトル特徴を反映する特徴量を取り出すための音声特徴抽出装置、404-1は入力画像に含まれる利用者の身長に代表される大局的な画像特徴量を抽出する画像特徴抽出装置、404-2は入力画像に含まれる利用者の顔色、顔の輪郭、目の大きさ、輝度分布等の局所的な画像特徴量を抽出する画像特徴抽出装置、405は予め登録されている利用者の音声の特徴量を蓄積している登録者音声特徴蓄積装置、406は音声特徴抽出装置403で取り出された現在の利用者の音声の特徴量と予め登録されている利用者の音声の特徴量を照合するための音声照合装置、407は予め登録されている利用者の画像の特徴量を蓄積している登録者画像特徴蓄積装置、408は画像特徴抽出装置404で取り出された現在の利用者の画像の特徴量と予め登録されている利用者の画像の特徴量を照合するための画像照合装置、409は画像照合装置408で照合された利用者と音声照合装置406で照合された利用者とが一致しているか否かを判定する一致判定装置である。音声特徴抽出装置403-1による大局的な音声の特徴量に基づいて登録者音声特徴蓄積装置405での音声候補を限定するように構成されている。また、画像特徴抽出装置404-1による大局的な画像の特徴量に基づいて登録者画像特徴蓄積装置407での画像候補を限定するように構成されている。

【0038】以上のように構成された実施の形態4の利用者照合装置について、以下にその動作を説明する。情報処理装置400の現在の利用者は利用に先立ち利用者の音声と顔画像を入力する。音声入力装置401から入力された現在の利用者の音声は音声特徴抽出装置403

-1、403-2に入力され、利用者の音声に含まれる個人性情報が抽出される。音声特徴抽出装置403-1では、利用者の個人性情報である音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差の大局的な特徴量を求める。音声特徴抽出装置403-2では音声パワースペクトラムの時系列情報を抽出する。これらの特徴量は公知の手法(例えばDFT分析、LPC分析、フィルタ分析等)を用いて容易に求めることが可能である。音声特徴抽出装置403-1で求めた大局的な音声特徴量は予め登録されている利用者の音声の特徴量を蓄積している登録者音声特徴蓄積装置405の照合すべき候補を限定するために用いられる。すなわち、現在の利用者の音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差の大局的な特徴量の類似している蓄積音声に登録者音声特徴蓄積装置405から選択され、音声照合装置406での照合に用いられる。入力された音声と限定された音声は音声照合装置406によって照合され、利用者が予め登録されているか否かが判断される。音声照合装置406での特徴量の照合は公知のHMM、DTW、LVQなどのパターン照合手法を用いて行うことができる。

【0039】一方、画像入力装置402によって入力された利用者の画像は画像特徴抽出装置403-1、403-2に入力され、利用者の画像に含まれる個人性情報が抽出される。画像特徴抽出装置403-1では、利用者の個人性情報である身長に代表されるような画像の大局的な特徴量を求め、登録者画像特徴蓄積装置407において照合すべき画像候補を限定する。画像特徴抽出装置403-2では顔画像の輝度情報、輝度情報の位置微分情報、顔画像の色情報を画像特徴量として抽出している。

【0040】画像照合装置408では、画像特徴抽出装置404-2で抽出された画像特徴量と登録者画像特徴蓄積装置407において限定された画像特徴量とを照合し、現在の利用者の画像が予め登録されている利用者の画像であるか否かを照合する。照合すべき登録蓄積画像の候補は本実施の形態では、画像特徴抽出装置404-1で求めた大局的な画像特徴量の類似した候補を登録者画像特徴蓄積装置407から限定している。一致判定装置409は音声照合装置406と画像照合装置408との照合結果を照合し、それらが一致したときのみ「1」を出力し、不一致のときは「0」を出力する。一致判定装置409が「1」を出力したときのみ、現在の利用者が登録されている利用者であると判断して、利用者が情報処理装置400を利用することを許可している。

【0041】以上のように本実施の形態によれば、利用者の音声の大局的な情報と画像の大局的な情報とを用いて照合候補を選択することによって、利用者照合時間を大きく削減することができ、利用者数が大きな情報処理装置を利用する場合に効率的な照合が可能となるとも

に、従来のパスワードをキーボードから入力するだけで  
は避けることのできない情報処理装置の不正利用を防ぐ  
ことができる。

【0042】（実施の形態5）以下、本発明の実施の形  
態5について図面を参照して説明する。

【0043】図5は本発明の実施の形態5における利用  
者照合装置の構成図を示すものである。この実施の形態  
5の構成は前述の実施の形態4（図4）の構成とほぼ同  
様であり、同様な情報処理装置500、音声入力装置5  
01、画像入力装置502、音声特徴量抽出装置503  
-1、503-2、画像特徴抽出装置504-1、50  
4-2、登録者音声特徴蓄積装置505、音声照合装置  
506、登録者画像特徴蓄積装置507、画像照合装置  
508、一致判定装置509を備えている。音声特徴抽  
出装置503-1は入力された現在の利用者の音声の平  
均的なピッチ等の大局的な特徴量を取り出し、音声特徴  
抽出装置503-2は入力された現在の利用者の音声の  
スペクトル特徴を反映する特徴量を取り出す。画像特徴  
抽出装置504-1は入力画像に含まれる利用者の身長  
に代表される大局的な画像特徴量を抽出し、画像特徴抽  
出装置504-2は入力画像に含まれる利用者の顔色、  
顔の輪郭、目の大きさ、輝度分布等の局所的な画像特徴  
量を抽出する。音声特徴抽出装置503-1による大  
局的な音声の特徴量に基づいて登録者音声特徴蓄積装置5  
05での音声候補を限定するように構成されている。また、  
画像特徴抽出装置504-1による大局的な画像の  
特徴量と音声照合装置506による音声照合結果とに基  
づいて登録者画像特徴蓄積装置507での画像候補を限  
定するように構成されている。

【0044】実施の形態4の場合に比べると、音声照合  
装置506の照合結果を用いて、照合すべき画像をさら  
に限定することによって、より効率的な照合を行うこと  
ができる。

【0045】（実施の形態6）以下、本発明の実施の形  
態6について図面を参照して説明する。

【0046】図6は本発明の実施の形態6における利用  
者照合装置の構成図を示すものである。この実施の形態  
6の構成は前述の実施の形態4（図4）の構成とほぼ同  
様であり、同様な情報処理装置600、音声入力装置6  
01、画像入力装置602、音声特徴量抽出装置603  
-1、603-2、画像特徴抽出装置604-1、60  
4-2、登録者音声特徴蓄積装置605、音声照合装置  
606、登録者画像特徴蓄積装置607、画像照合装置  
608、一致判定装置609を備えている。音声特徴抽  
出装置603-1は入力された現在の利用者の音声の平  
均的なピッチ等の大局的な特徴量を取り出し、音声特徴  
抽出装置603-2は入力された現在の利用者の音声の  
スペクトル特徴を反映する特徴量を取り出す。画像特徴  
抽出装置604-1は入力画像に含まれる利用者の身長  
に代表される大局的な画像特徴量を抽出し、画像特徴抽

出装置604-2は入力画像に含まれる利用者の顔色、  
顔の輪郭、目の大きさ、輝度分布等の局所的な画像特徴  
量を抽出する。画像特徴抽出装置604-1による大  
局的な画像の特徴量に基づいて登録者画像特徴蓄積装置6  
07での画像候補を限定するように構成されている。ま  
た、音声特徴抽出装置603-1による大局的な音声の  
特徴量および画像照合装置608による画像照合結果と  
に基づいて登録者音声特徴蓄積装置605での音声候補  
を限定するように構成されている。

【0047】実施の形態4の場合に比べると、画像照合  
装置608の照合結果を用いて、照合すべき音声をさら  
に限定することによって、より効率的な照合を行うこと  
ができる。

【0048】（実施の形態7）以下、本発明の実施の形  
態7について図面を参照して説明する。

【0049】図7は本発明の実施の形態7における利用  
者照合装置の構成図を示すものである。図7において、  
700は情報処理装置、701は情報処理装置700に  
現在の利用者の音声を入力するための音声入力装置、7  
02は情報処理装置700に現在の利用者の顔画像を入  
力するための画像入力装置、703-1は音声入力装置  
701で入力された現在の利用者の音声の平均的なピ  
ッチ等の大局的な特徴量を取り出すための音声特徴抽出  
装置、703-2は音声入力装置701で入力された現在  
の利用者の音声のスペクトル特徴を反映する特徴量を取  
り出すための音声特徴抽出装置、704は画像入力装置  
702で入力された現在の利用者の画像の特徴量を取  
り出すための画像特徴抽出装置、705は予め登録され  
ている利用者の音声の特徴量を蓄積している登録者音声特  
徴蓄積装置、706は音声特徴抽出装置703で取り出  
された現在の利用者の音声の特徴量と予め登録されてい  
る利用者の音声の特徴量を照合するための音声照合装置  
である。音声特徴抽出装置703-1による大局的な音  
声の特徴量に基づいて登録者音声特徴蓄積装置705で  
の音声候補を限定するように構成されている。707は  
予め登録されている利用者の画像の特徴量を蓄積してい  
る登録者画像特徴蓄積装置、708は画像特徴抽出装置  
704で取り出された現在の利用者の画像の特徴量と予  
め登録されている利用者の画像の特徴量を照合するた  
めの画像照合装置である。音声照合装置706によるN個  
の音声候補に基づいて登録者画像特徴蓄積装置707で  
の画像候補を限定するように構成されている。709は  
画像照合装置708で照合した結果の信頼度Gを求める  
ための画像信頼度評価装置、710は音声照合装置70  
6で照合した結果の信頼度Sを求めるための音声信頼度  
評価装置、711は画像信頼度評価装置709で求めら  
れた画像から評価した現在の利用者が登録利用者である  
ことを判断するための信頼度Gと、音声信頼度評価装置  
710で求められた音声から評価した現在の利用者が登  
録利用者であることを判断するための信頼度Sを総合し

て現在の利用者が情報処理装置 700 を利用することを許可するか否かを判断するための信頼度評価装置である。

【0050】以上のように構成された実施の形態 7 の利用者照合装置について、以下にその動作を説明する。情報処理装置 700 の現在の利用者は利用に先立ち利用者の音声と顔画像を入力する。音声入力装置 701 から入力された音声は音声特徴抽出装置 703-1, 703-2 に入力され利用者の音声に含まれる個人性情報が抽出される。音声特徴抽出装置 703-1 では、音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差の個人性情報の大局的な特徴量を求める。音声特徴抽出装置 703-2 では、音声パワースペクトラムの時系列情報を抽出する。これらの特徴量は公知の手法（例えば DFT 分析、LPC 分析、フィルタ分析等）を

$$S = 1 - S_{\min} / S_{\min 2} \quad \dots \dots \dots (1)$$

で定義している。式 (1) で  $S_{\min}$  は入力音声と最も良く照合した蓄積音声との照合結果である距離であり、 $S_{\min 2}$  は次に良く照合した蓄積音声との距離である。信頼度  $S$  が 1 に近いほど入力音声と蓄積音声との音声との一致度が高く信頼度が高いことになる。

【0051】一方、画像入力装置 702 によって入力された現在の利用者の画像は画像特徴抽出装置 704 に入力され、利用者の画像に含まれる個人性情報を抽出する。画像に含まれる個人性情報としては、顔画像の輝度情報、輝度情報の位置微分情報、顔画像の色情報を用いる。現在の利用者の画像が予め情報処理装置 700 に登録されている利用者の画像であるか否かは、画像特徴抽出装置 704 で取り出した特徴量と登録者画像特徴蓄積

$$G = 1 - G_{\min} / G_{\min 2} \quad \dots \dots \dots (2)$$

で定義している。式 (2) で  $G_{\min}$  は入力画像と最も良く照合した蓄積画像との照合結果である距離であり、 $G_{\min 2}$  は次に良く照合した蓄積画像との距離である。本実施の形態の照合距離は、照合する画像間の対応する画素間の輝度の差を使って求めている。信頼度  $G$  が 1 に近いほど入力画像と蓄積画像との一致度が高く信頼度が高いことになる。

【0052】信頼度評価装置 711 では音声信頼度評価装置 710 で求めた信頼度  $S$  と画像信頼度評価装置 709 で求めた信頼度  $G$  のどちらともが予め設定している基準値よりも高く、かつ音声照合装置 706 と画像照合装置 708 による音声で判定した利用者と画像で判定した利用者が一致したときにのみ現在の利用者が登録された利用者であるとして、情報処理装置 700 の利用を許可する。

【0053】以上のように本実施の形態によれば、入力された音声と画像を用いて利用者を判定すると同時に、音声と画像の信頼度を考慮することでより信頼度の高い確実な利用者照合判定を行うことができる。なお、音声で判定した利用者と画像で判定した利用者が一致しては

用いて容易に求めることが可能である。音声特徴抽出装置 703-1 で求められた音声の特徴量は登録者音声特徴蓄積装置 705 に入力され、音声照合装置 706 で照合に用いる候補データを選出する。音声照合装置 706 では登録者音声特徴蓄積装置 705 からの候補データと現在の利用者について、音声特徴抽出装置 703-2 で抽出された特徴量を用いて照合を行い、その結果を登録者画像特徴蓄積装置 707 と音声信頼度評価装置 710 に出力する。音声照合装置 706 での特徴量の照合は公知の HMM、DTW、LVQ などのパターン照合手法を用いて行うことができる。これらの照合の結果、現時点で入力された音声と予め蓄積されている音声データとどれくらい似ているかが音声信頼度評価装置 710 で評価される。本実施の形態では、音声信頼度評価装置 710 での信頼度  $S$  の評価は具体的には、

装置 707 の特徴量とを画像照合装置 708 で照合する。照合に用いられる蓄積画像は、音声信号装置 706 から出力された現在の利用者に最も近い登録利用者から  $N$  番目 ( $N$  は正の整数) までの候補を用いる。現在の利用者の画像と予め情報処理装置 700 に登録されている利用者の照合は入力画像の輝度情報等を使う公知の手法で行うことが可能である。画像照合装置 708 で求められる現在の利用者の入力画像と蓄積画像との類似度は画像信頼度評価装置 709 に入力される。画像信頼度評価装置 709 では入力された照合結果から信頼度を求める。画像信頼度評価装置 709 での信頼度  $G$  の評価は具体的には、

いるが、信頼度が低い場合には、さらに現在の利用者によりキーボードからパスワードを入力させることによって情報処理装置 700 の利用を許可することもできる。また、信頼度評価装置 711 の判定基準としては、音声信頼度評価装置 710 で求めた信頼度  $S$  と画像信頼度評価装置 709 で求めた信頼度  $G$  のどちらともが予め設定している基準値を超えたことを基準とする代わりに、信頼度  $S$  または信頼度  $G$  が 1 に近いときには情報処理装置 700 の利用を許可するようにすることもできる。

【0054】（実施の形態 8）以下、本発明の実施の形態 8 について図面を参照して説明する。

【0055】図 8 は本発明の実施の形態 8 における利用者照合装置の構成図を示すものである。図 8 において、800 は情報処理装置、801 は情報処理装置 800 に現在の利用者の音声を入力するための音声入力装置、802 は情報処理装置 800 に現在の利用者の顔画像を入力するための画像入力装置、803-1 は音声入力装置 801 で入力された現在の利用者の音声の平均的なピッチ等の大局的な特徴量を取り出すための音声特徴抽出装置、803-2 は音声入力装置 801 で入力された現在

の利用者の音声のスペクトル特徴を反映する特徴量を取り出すための音声特徴抽出装置、804は画像入力装置802で入力された現在の利用者の画像の特徴量を取り出すための画像特徴抽出装置、805は予め登録されている利用者の音声の特徴量を蓄積している登録者音声特徴蓄積装置、806は音声特徴抽出装置803で取り出された現在の利用者の音声の特徴量と予め登録されている利用者の音声の特徴量を照合するための音声照合装置である。音声特徴抽出装置803-1による大局的な音声の特徴量に基づいて登録者音声特徴蓄積装置805での音声候補を限定するように構成されている。807は予め登録されている利用者の画像の特徴量を蓄積している登録者画像特徴蓄積装置、808は画像特徴抽出装置804で取り出された現在の利用者の画像の特徴量と予め登録されている利用者の画像の特徴量を照合するための画像照合装置、809は画像照合装置808で照合した結果の信頼度を求めるための画像信頼度評価装置である。画像信頼度評価装置809によって得た信頼度Gに基づいて登録者音声特徴蓄積装置805でN個の音声候補を限定するように構成されている。810は音声照合装置806で照合した結果の信頼度Sを求めるための音声信頼度評価装置、811は画像信頼度評価装置809で求められた画像から評価した現在の利用者が登録利用者であることを判断するための信頼度Gと、音声信頼度評価装置810で求められた音声から評価した現在の利用者が登録利用者であることを判断するための信頼度Sを総合して現在の利用者が情報処理装置800を利用することを許可するか否かを判断するための信頼度評価装置である。

【0056】以上のように構成された実施の形態8の利用者照合装置について、以下にその動作を説明する。情報処理装置800の現在の利用者は利用に先立ち利用者の音声と顔画像を入力する。画像入力装置802によって入力された利用者の画像は、画像特徴抽出装置804に入力され利用者の画像に含まれる個人性情報が抽出される。画像の個人性情報としては、顔画像の輝度情報、輝度情報の位置微分情報、顔画像の色情報を用いる。画像照合装置808では登録者画像特徴蓄積装置807に蓄積されている画像特徴量と画像特徴抽出装置804で抽出された画像特徴量とを照合し、現在の利用者にもっとも近い登録利用者からN番目(Nは正の整数)までの候補を画像信頼度評価装置809に入力する。画像信頼度評価装置809では信頼度に応じて照合候補数Nを求め、選択されたN個の候補を登録者音声特徴蓄積装置805に出力する。すなわち信頼度が高いときには少ない照合候補を、信頼度が低いときには多くの照合候補を登録者音声特徴蓄積装置805に出力する。登録者音声特徴蓄積装置805では照合すべき音声候補を限定する。

【0057】画像信頼度評価装置809での信頼度Gの評価は具体的には、 $G = 1 - G_{\min} / G_{\min 2}$ で定義して

おり、 $G_{\min}$ は入力画像と最も良く照合した蓄積画像との距離であり、 $G_{\min 2}$ は次に良く照合した蓄積画像との照合結果である距離である。本実施の形態の照合距離は、照合する画像間の対応する画素間の輝度の差を使って求めている。信頼度Gが1に近いほど入力画像と蓄積画像との一致度が高く信頼度が高いため、登録者音声特徴蓄積装置805に出力される候補数は少なくよい。

【0058】一方、音声入力装置801から入力された現在の利用者の音声は音声特徴抽出装置803-1、803-2に入力されて、利用者の音声に含まれる個人性情報が抽出される。音声特徴抽出装置803-1では、音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差の個人性情報の大局的な特徴量を求める。音声特徴抽出装置803-2では、音声パワースペクトラムの時系列情報等を求める。これらの特徴量は公知の手法(例えばD-T分析、LPC分析、フィルタ分析等)を用いて容易に求めることが可能である。音声特徴抽出装置803-1、803-2で求められた音声の特徴量は登録者音声特徴蓄積装置805の情報と音声照合装置806で照合される。ここで照合に用いられる蓄積音声は、画像信頼度評価装置809から画像信頼度に応じて出力されたN個の利用者の音声候補と、音声特徴抽出装置803-1で求めた大局的な音声特徴量で限定されたデータを用いている。音声照合装置806での特徴量の照合は公知のHMM、DTW、LVQなどのパターン照合手法を用いて行うことができる。これらの照合の結果、現時点で入力された音声が予め蓄積されている音声データとどれくらい似ているかが音声信頼度評価装置810で評価される。本実施の形態では、音声信頼度評価装置810での信頼度Sの評価は具体的には、 $S = 1 - S_{\min} / S_{\min 2}$ で定義しており、 $S_{\min}$ は入力音声と最も良く照合した蓄積音声との距離であり、 $S_{\min 2}$ は次に良く照合した蓄積音声との距離である。信頼度Sが1に近いほど入力音声と蓄積音声との一致度が高く信頼度が高いことになる。

【0059】信頼度評価装置811では音声信頼度評価装置810で求めた信頼度Sと画像信頼度評価装置809で求めた信頼度Gに基づいて判定を行う。ここでの判定は、信頼度S、Gに基づいて予め設定している図9に示したような判定基準に基づいて行っている。音声信頼度評価装置810と画像信頼度評価装置809で求めた信頼度S、Gがそれぞれ予め設定した値よりも高く図9で網点を施した領域に入ったときに登録された利用者であるとして、情報処理装置800の利用を許可する。この判定基準作成は音声信頼度評価装置810と画像信頼度評価装置809で収集した信頼度からテーブルとして作成することができるほか、公知のニューラルネットワークやファジー等の手法を用いて容易に行うことができる。

【0060】以上のように本実施の形態によれば、入力

された画像と音声を用いて利用者を判定すると同時に、音声と画像の信頼度を考慮することでより信頼度の高い確実な利用者照合判定を行うことができる。なお、音声で判定した利用者と画像で判定した利用者が一致しているが、信頼度が低い場合には、現在の利用者にキーボードからパスワードを入力させることによって情報処理装置 800 の利用を許可することもできる。

【0061】（実施の形態 9）以下、本発明の実施の形態 9 について図面を参照して説明する。

【0062】図 10 は本発明の実施の形態 9 における利用者照合装置の構成図を示すものである。図 10 において、1000 は情報処理装置、1001 は情報処理装置 1000 に現在の利用者の音声を入力するための音声入力装置、1002 は情報処理装置 1000 に現在の利用者の顔画像を入力するための画像入力装置、1003-1 は音声入力装置 1001 で入力された現在の利用者の音声の平均的なピッチ等の大局的な特徴量を取り出すための音声特徴抽出装置、1003-2 は音声入力装置 1001 で入力された現在の利用者の音声のスペクトル特徴を反映する特徴量を取り出すための音声特徴抽出装置、1004 は画像入力装置 1002 で入力された現在の利用者の顔画像の特徴量を取り出すための画像特徴抽出装置、1005 は予め登録されている利用者の音声の特徴量を蓄積している登録者音声特徴蓄積装置、1006 は音声特徴抽出装置 1003 で取り出された現在の利用者の音声の特徴量と予め登録されている利用者の音声の特徴量を照合するための音声照合装置、1007 は予め登録されている利用者の画像の特徴量を蓄積している登録者画像特徴蓄積装置、1008 は画像特徴抽出装置 1004 で取り出された現在の利用者の画像の特徴量と予め登録されている利用者の画像の特徴量を照合するための画像照合装置、1009 は画像照合装置 1008 で照合した結果の信頼度  $G$  を求めるための画像信頼度評価装置、1010 は音声照合装置 1006 で照合した結果の信頼度  $S$  を求めるための音声信頼度評価装置、1011 は画像信頼度評価装置 1009 で求められた画像から評価した現在の利用者が登録利用者であることを判断するための信頼度  $G$  と音声信頼度評価装置 1010 で求められた音声から評価した現在の利用者が登録利用者であることを判断するための信頼度  $S$  を総合して現在の利用者が情報処理装置 1000 を利用することを許可するか否かを判断するための信頼度評価装置である。音声特徴抽出装置 1003-1 による大局的な音声の特徴量に基づいて登録者音声特徴蓄積装置 1005 での音声候補を限定するように構成されている。また、画像信頼度評価装置 1009 によって得た信頼度  $G$  に基づいて登録者音声特徴蓄積装置 1005 で  $N$  個の音声候補を限定するように構成されている。

【0063】以上のように構成された実施の形態 9 の利用者照合装置について、以下にその動作を説明する。情

報処理装置 1000 の現在の利用者は利用に先立ち利用者の音声と顔画像を入力する。画像入力装置 1002 によって入力された現在の利用者の画像は、画像特徴抽出装置 1004 に入力され利用者の画像に含まれる個人性情報が抽出される。画像の個人性情報としては、顔画像の輝度情報、輝度情報の位置微分情報、顔画像の色情報を用いる。画像照合装置 1008 では登録者画像特徴蓄積装置 1007 に蓄積されている画像特徴量と特徴抽出装置 1004 からの画像特徴量とを照合し、現在の利用者 10 者に最も近い登録利用者から  $N$  番目 ( $N$  は正の整数) までの候補を画像信頼度評価装置 1009 に入力する。画像信頼度評価装置 1009 では信頼度の最も高い候補の信頼度に応じて、 $N$  個の候補を求め、その候補情報を登録者音声特徴蓄積装置 1005 に出力する。すなわち、類似度が高いときには少ない候補を、類似度が低いときには多くの候補を登録者音声特徴蓄積装置 1005 に出力する。画像信頼度評価装置 1009 での信頼度  $G$  の評価は具体的には、 $G = \frac{1}{G_{\min} - G_{\min 2}}$  で定義しており、 $G_{\min}$  は入力画像と最も良く照合した蓄積画像との照合結果の距離であり、 $G_{\min 2}$  は次に良く照合した蓄積画像との距離である。本実施の形態の照合距離は、照合する画像間の対応する要素間の輝度の差を使って求めている。信頼度  $G$  が 1 に近いほど入力画像と蓄積画像との一致度が高く信頼度が高いことになる。本実施の形態ではここで求めた画像信頼度  $G$  が予め設定した値よりも高かったときには、現在の利用者が登録されている利用者であると認め、情報処理装置 1000 の利用を許可する。

【0064】一方、画像信頼度  $G$  が予め設定した値よりも低かったときには音声を用いた照合を行うために以下の処理を行う。音声入力装置 1001 より入力された現在の利用者の音声は音声特徴抽出装置 1003-1、1003-2 に入力されて、利用者の音声に含まれる個人性情報を抽出する。音声特徴抽出装置 1003-1 では、音声のピッチの平均値、ピッチの標準偏差、平均音声パワーとその標準偏差の個人性情報の大局的な特徴量を求める。音声特徴抽出装置 1003-2 では、音声パワースペクトラムの時系列情報等を求める。これらの特徴量は公知の手法（例えば DFT 分析、LPC 分析、フィルタ分析等）を用いて容易に求めることが可能である。音声特徴抽出装置 1003-1、1003-2 で求められた音声の特徴量は登録者音声特徴蓄積装置 1005 の特徴量と音声照合装置 1006 で照合される。ここで照合に用いられる蓄積音声は、画像信頼度評価装置 1009 から画像信頼度  $G$  に応じて出力された  $N$  個の利用者の候補と、音声特徴抽出装置 1003-1 で求めた大局的な音声特徴量で限定されたデータを用いている。音声照合装置 1006 での特徴量の照合は公知の HMM、DTW、LVQ などのパターン照合手法を用いて行うことができる。これらの照合の結果、現時点で入力された

音声が予め蓄積されている音声データとどれくらい似ているかが音声信頼度評価装置 1010 で評価される。本実施の形態では、音声信頼度評価装置 1010 での信頼度  $S$  の評価は具体的には、 $S = 1 - S_{\min} / S_{\min 2}$  で定義しており、 $S_{\min}$  は入力音声と最も良く照合した蓄積音声との距離であり、 $S_{\min 2}$  は次に良く照合した蓄積音声との距離である。信頼度  $S$  が 1 に近いほど入力音声と蓄積音声との一致度が高く信頼度が高いことになる。

【0065】信頼度評価装置 1011 では画像信頼度評価装置 1009 で求めた画像信頼度  $G$  と音声信頼度評価装置 1010 で求めた音声信頼度  $S$  がともに予め設定した基準値よりも高く、かつ音声で判定した利用者と画像で判定した利用者が一致したときにのみ現在の利用者が登録された利用者であるとして、情報処理装置 1000 の利用を許可する。

【0066】以上のように本実施の形態によれば、まず画像を用いて利用者を照合し、そのときの画像信頼度が設定値よりも高かった場合には音声を使った照合は行わずに情報処理装置 1000 の利用を許可し、画像信頼度が設定値よりも低かった場合には音声を用いた照合も行うことで、より効率的で確かな利用者照合判定を行うことができる。

#### 【0067】

【発明の効果】以上説明したように、本発明によれば、利用者に特有な音声と顔画像の情報を用いて利用者の照合を行うことによって情報処理装置の利用者を確実に制限することができ、従来のパスワード（暗唱番号）をキーボードから入力するだけでは避けることのできない情報処理装置の不正利用を防ぐことができ、その実用的価値には大なるものがある。

#### 【図面の簡単な説明】

【図 1】本発明の実施の形態 1 における利用者照合装置の構成図である。

【図 2】本発明の実施の形態 2 における利用者照合装置の構成図である。

【図 3】本発明の実施の形態 3 における利用者照合装置の構成図である。

【図 4】本発明の実施の形態 4 における利用者照合装置の構成図である。

【図 5】本発明の実施の形態 5 における利用者照合装置

の構成図である。

【図 6】本発明の実施の形態 6 における利用者照合装置の構成図である。

【図 7】本発明の実施の形態 7 における利用者照合装置の構成図である。

【図 8】本発明の実施の形態 8 における利用者照合装置の構成図である。

【図 9】本発明の実施の形態 8 における信頼度評価装置の説明図である。

10 【図 10】本発明の実施の形態 9 における利用者照合装置の構成図である。

【図 11】従来の利用者照合装置の構成図である。

#### 【符号の説明】

100, 200, 300, 400, 500, 600, 700, 800, 1000 …… 情報処理装置

101, 201, 301, 401, 501, 601, 701, 801, 1001 …… 音声入力装置

102, 202, 302, 402, 502, 602, 702, 802, 1002 …… 画像入力装置

20 103, 203-1, 203-2, 303-1, 303-2, 403-1, 403-2, 503-1, 503-2, 603-1, 603-2, 703-1, 703-2, 803-1, 803-2, 1003-1, 1003-2 …… 音声特徴抽出装置

104, 204, 304, 404-1, 404-2, 504-1, 504-2, 604-1, 604-2, 704, 804, 1004 …… 画像特徴抽出装置

105, 205, 305, 405, 505, 605, 705, 805, 1005 …… 登録者音声特徴蓄積装置

30 106, 206, 306, 406, 506, 606, 706, 806, 1006 …… 音声照合装置

107, 207, 307, 407, 507, 607, 707, 807, 1007 …… 登録者画像特徴蓄積装置

108, 208, 308, 408, 508, 608, 708, 808, 1008 …… 画像照合装置

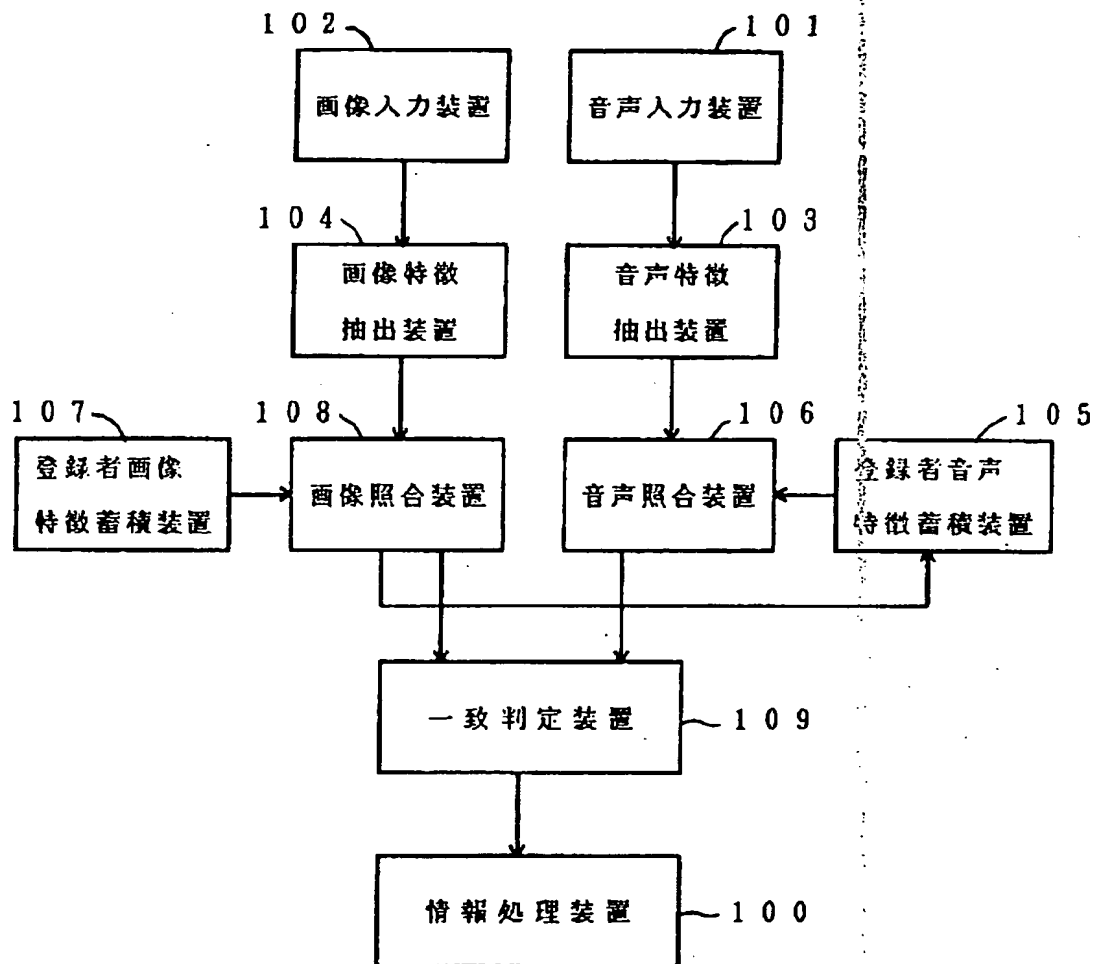
109, 209, 309, 409, 509, 609 …… 一致判定装置

709, 809, 1009 …… 画像信頼度評価装置

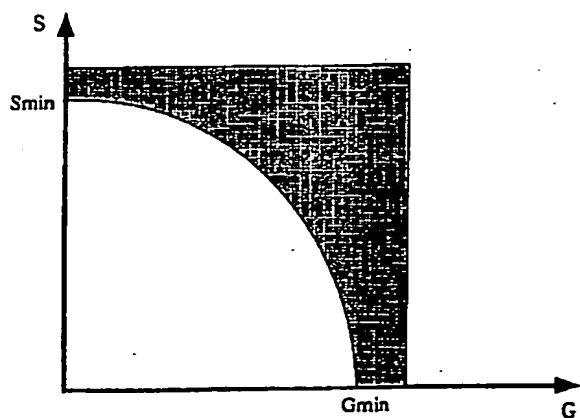
710, 810, 1010 …… 音声信頼度評価装置

40 711, 811, 1011 …… 信頼度評価装置

【図1】

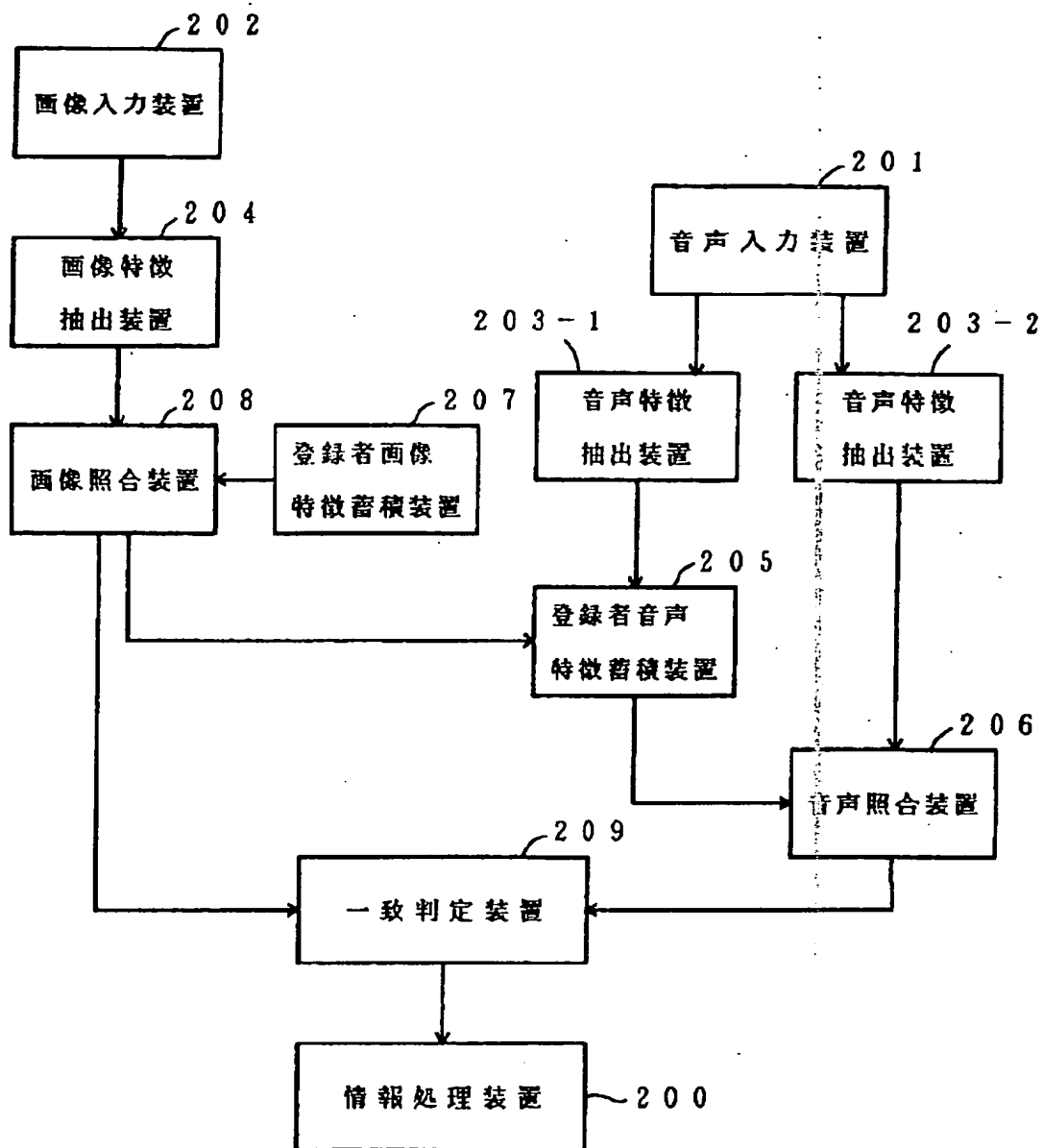


【図9】

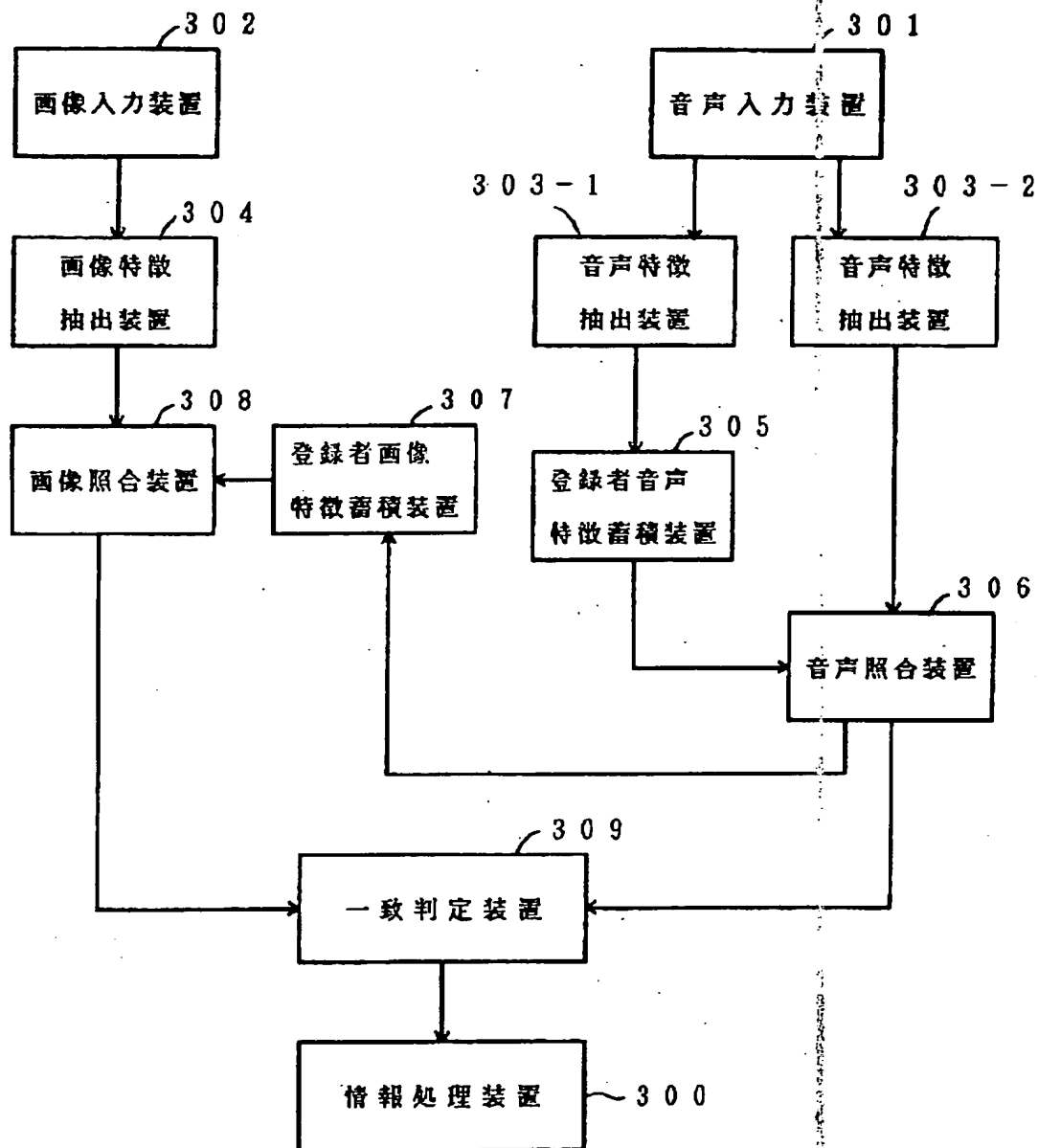




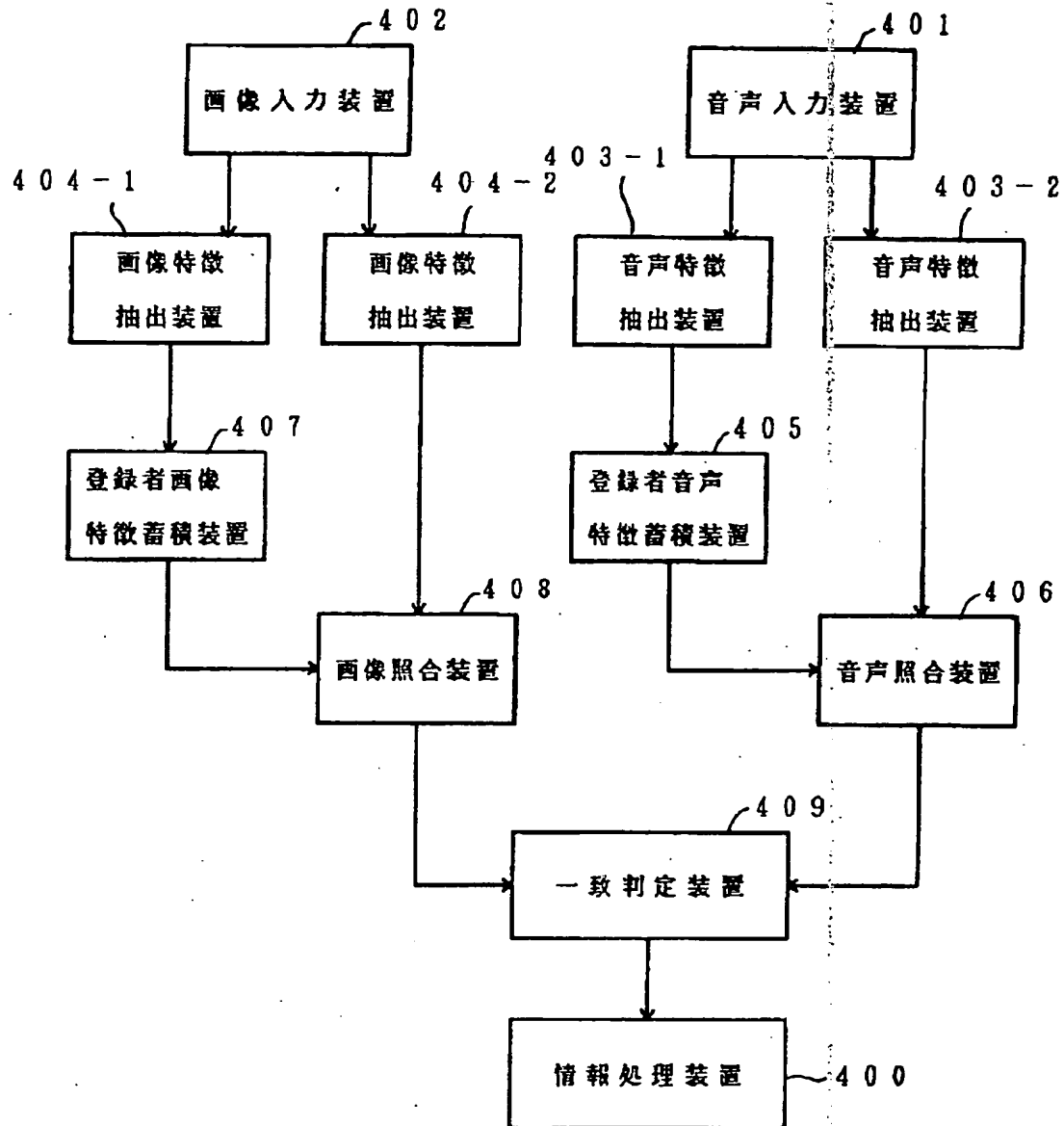
【図2】



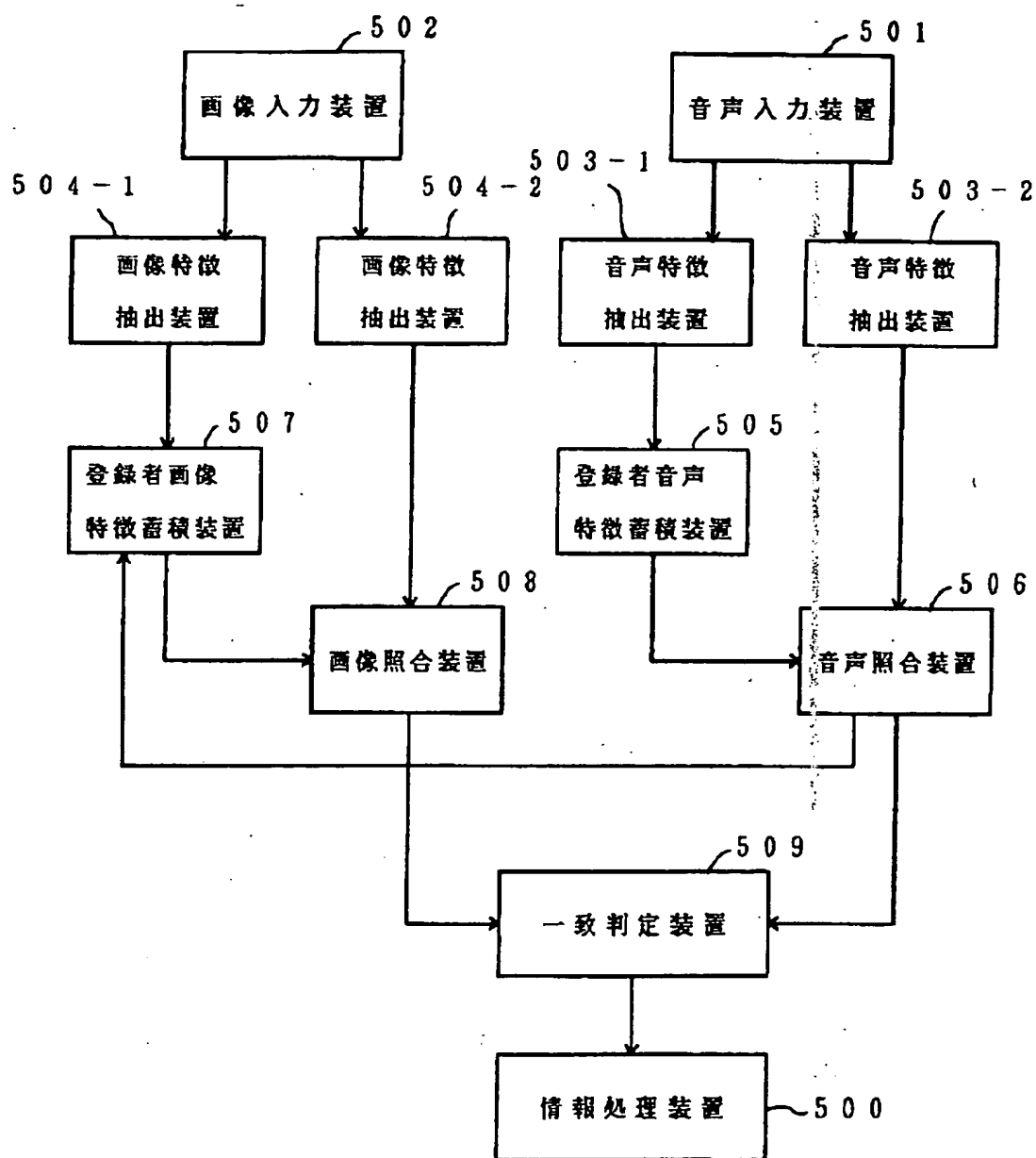
【図3】



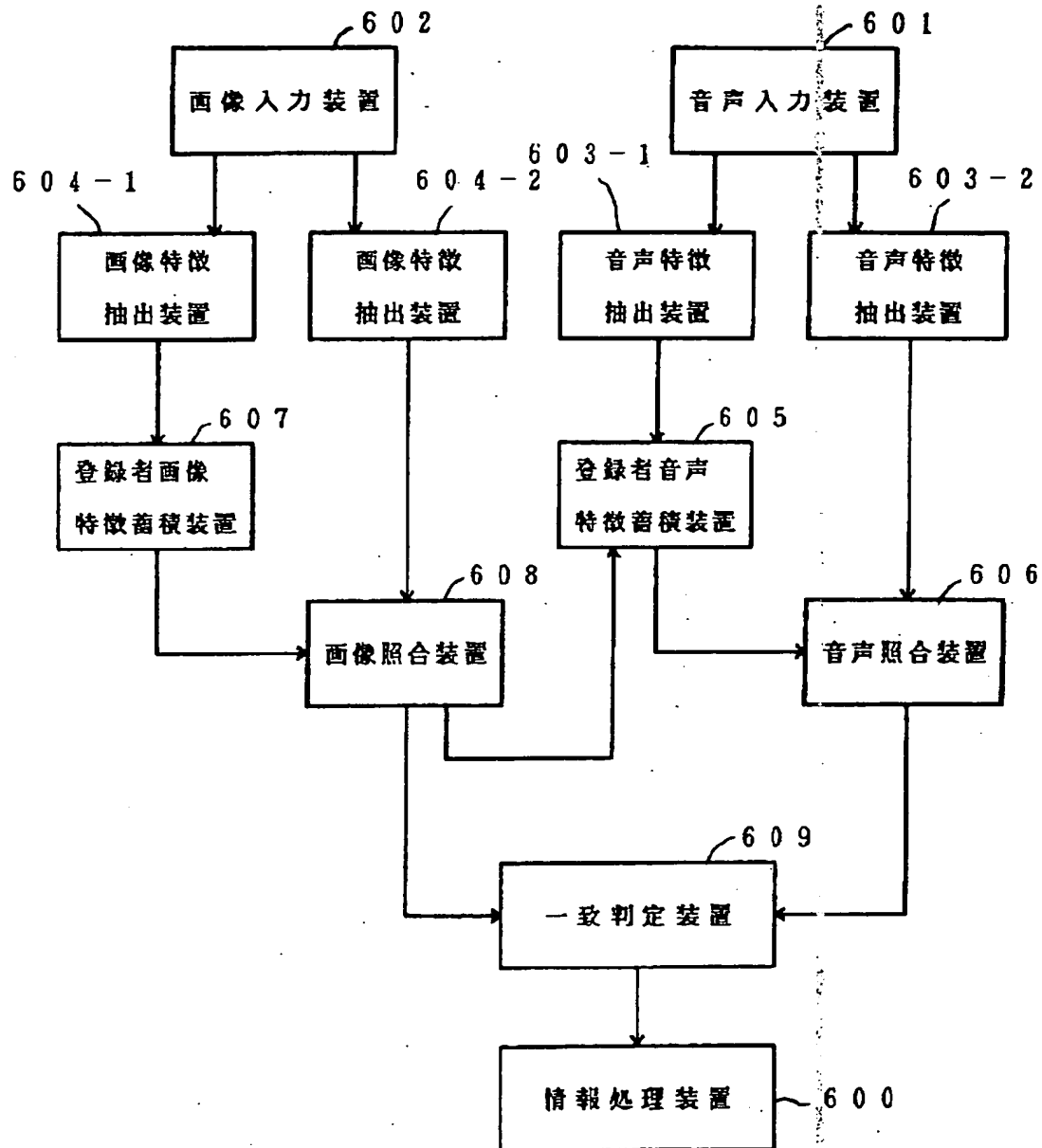
【図4】



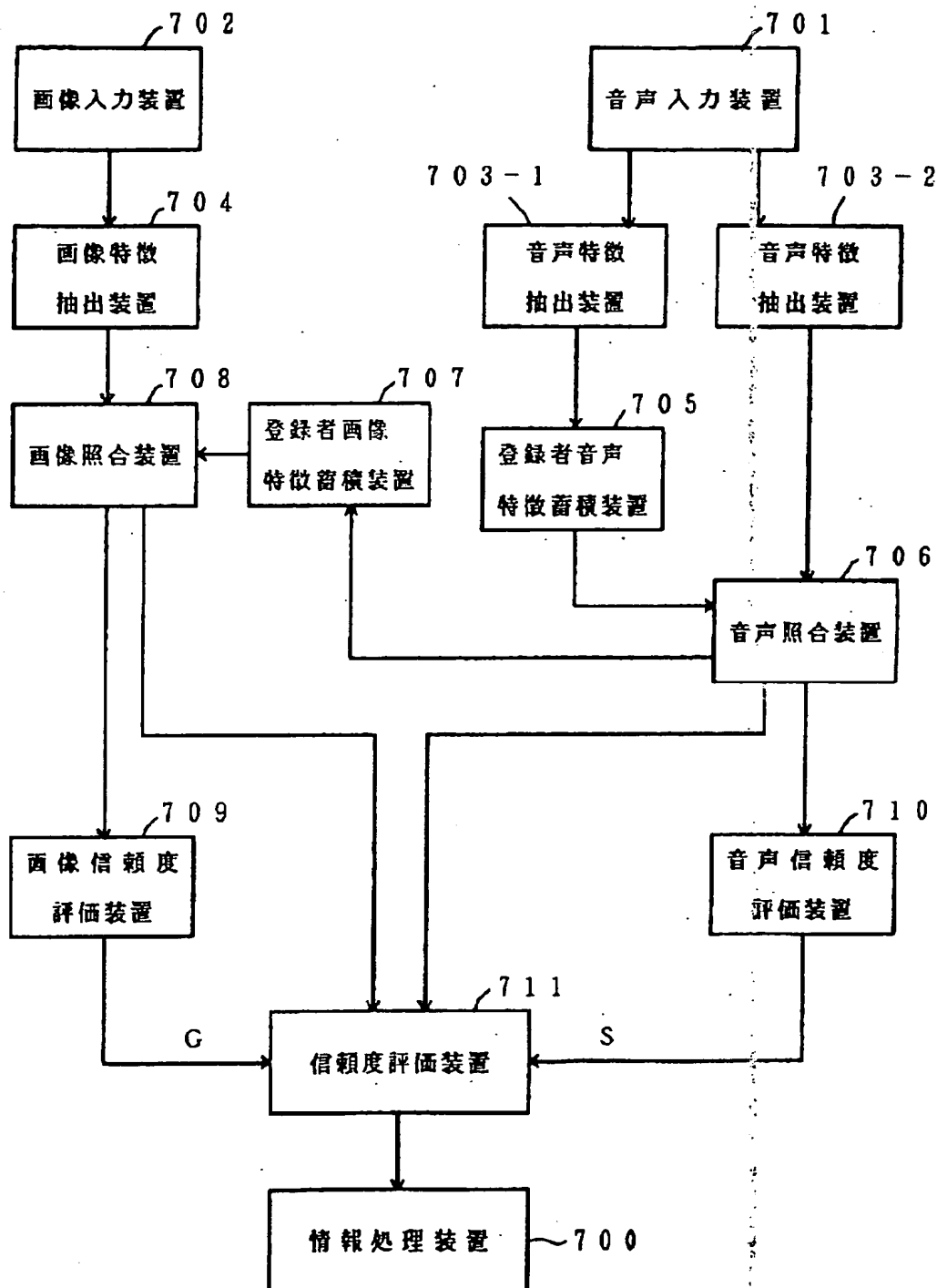
【図5】



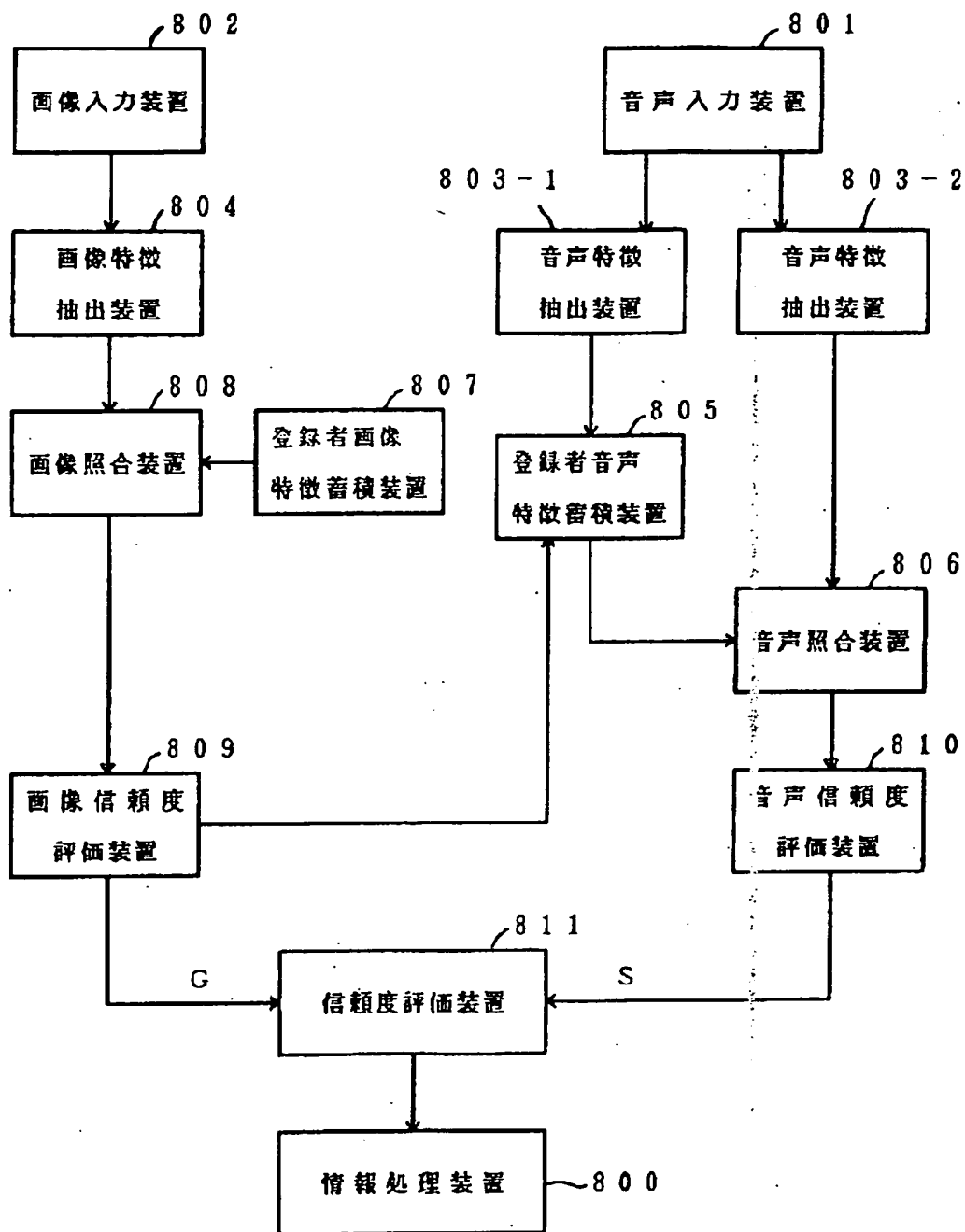
【図6】



【図7】

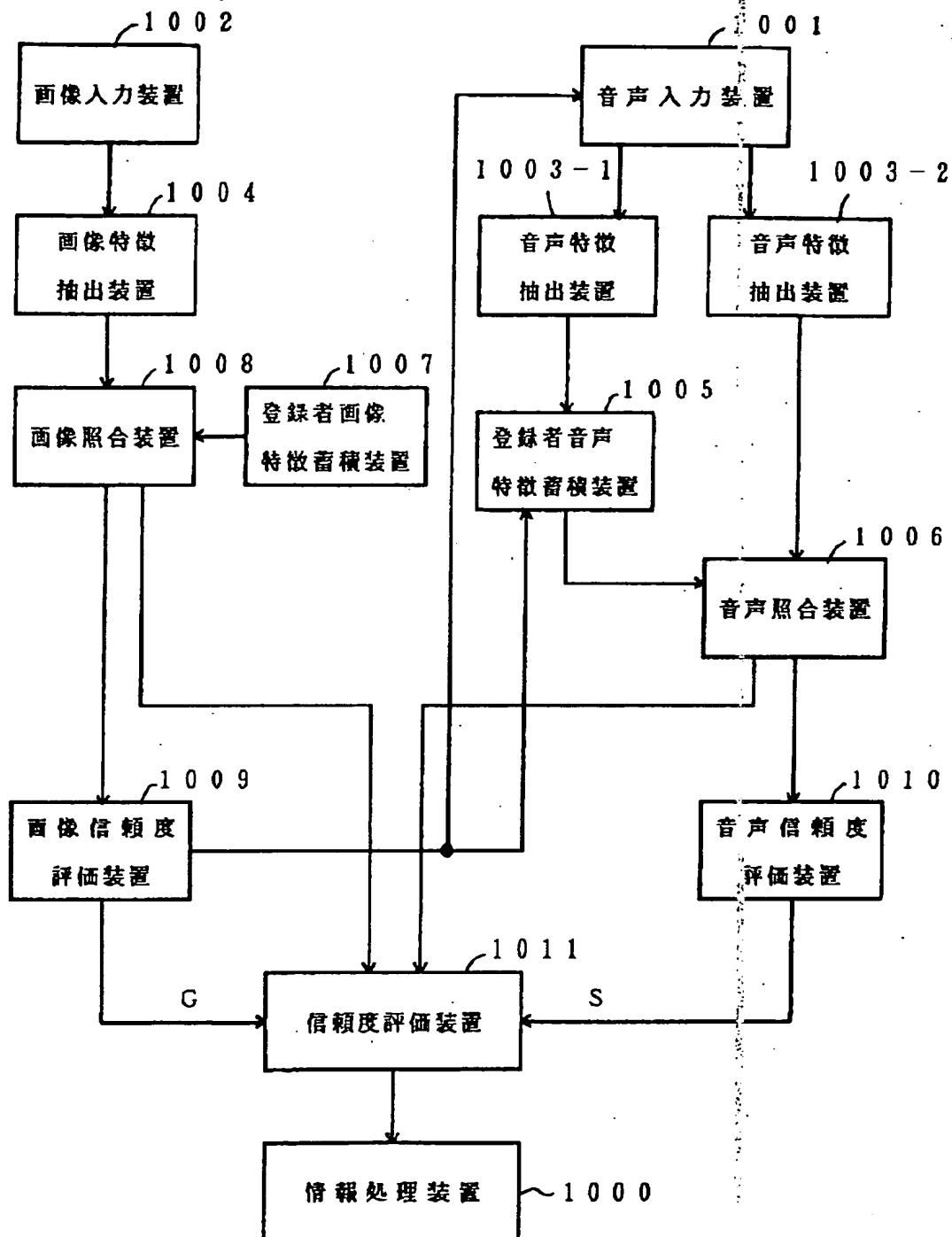


【図8】





【図10】



【図11】

